■12 群（電子情報通信基礎）－ 3 編（統計・確率）

7 章 統計的推測の基礎

（執筆者：青嶋　誠）[2009 年 2 月 受領]

■概要■

母集団分布が未知の母数に依存すると考える場合に，その分布から得られた標本の値に基づいて母数に何らかの近似値を与えることを**推定**という．推定は**点推定**と**領域推定**に大別される．点推定は母数の近似値を 1 点で与え，領域推定は母数の近似値を存在範囲で与える．特に，その範囲が区間になるとき，領域推定を**区間推定**という．

母集団に関する何らかの結論を導き出したいときに，母数に関する仮説を設けて，これが正しいか否かを標本に基づいて判定することを**仮説検定**という．母集団分布から得られた無作為標本 $\boldsymbol{X}$ に基づいて適当な領域 $R_{\boldsymbol{X}}$ を定めて，$\boldsymbol{X}$ の実現値 $\boldsymbol{x}$ について，$\boldsymbol{x} \in R_{\boldsymbol{X}}$ のとき仮説を棄却して，$\boldsymbol{x} \notin R_{\boldsymbol{X}}$ のとき仮説を受容する．ここで，$R_{\boldsymbol{X}}$ を $\boldsymbol{X}$ に基づく**棄却域**という．

本章では，統計的推測の根幹をなす推定と検定について，その基礎となる概念を解説する．特に，区間推定と仮説検定を構築する際に，その基となる母数に何らかの意味で良い推定量を求めるための方法を解説する．

【本章の構成】

本章では，点推定（7-1 節），モーメント法（7-2 節），最尤法（7-3 節），十分性（7-4 節），一様最小分散不偏推定（7-5 節），区間推定（7-6 節），仮説検定（7-7 節），統計モデル（7-8 節）に関する基礎知識について述べる．

■12 群 - 3 編 - 7 章

7-1 点推定

（point estimation）

（執筆者：青嶋　誠）［2009 年 2 月 受領］

母集団分布が p.d.f. または p.m.f. $p(x, \theta)$ $(\theta \in \Theta)$ を持つとする．ここで，θ を**母数**または**パラメータ**，Θ を**母数空間**といい，$\Theta \subset \boldsymbol{R}^k$ とする．いま，この母集団分布からの大きさ n の無作為標本を $X_1, \cdots, X_n$ とする．すなわち，確率変数 $X_1, \cdots, X_n$ は互いに独立にいずれもその分布に従うとする．このとき，$\boldsymbol{X} = (X_1, \cdots, X_n)$ の関数 $\hat{\theta}(\boldsymbol{X})$ が Θ 上の値をとれば，それを θ の**推定量**（estimator）という．特に，$\boldsymbol{X}$ が実現値 $\boldsymbol{x} = (x_1, \cdots, x_n)$ をとるとき，$\hat{\theta}(\boldsymbol{x})$ を θ の**推定値**（estimate）という．各 θ_i の推定量を $\hat{\theta}_i = \hat{\theta}_i(\boldsymbol{X})$ として $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ が $(\theta_1, \cdots, \theta_k)$ の推定量になる．また，θ の関数 $g(\theta)$ について，$\boldsymbol{X}$ の関数 $\hat{g}(\boldsymbol{X})$ が $g(\Theta) = \{g(\theta) | \theta \in \Theta\}$ 上の値をとるとき，それを $g(\theta)$ の推定量という．なお，θ が多（次元）母数で $\theta = (\xi, \eta)$ と表され，ξ が関心のある部分で，η は無関心部分とするとき，η を**局外母数**（nuisance parameter または**攪乱母数**）という．一般に，母数 θ の推定量を求める古典的な方法として，**モーメント法**と**最尤法**がよく知られている．

■12 群 - 3 編 - 7 章

7-2 モーメント法

（the method of moments）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

X_1 の原点周りの r 次のモーメントを $\mu'_{r,\theta} = E_\theta[X_1^r]$ $(r = 1, 2, \cdots)$ とする．ただし，$E_\theta[\cdot]$ は母数 θ を持つ分布に関する期待値を表す．また，その分布からの無作為標本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ に基づく r 次の標本モーメントを $M'_r = (1/n)\sum_{i=1}^n X_i^r$ $(r = 1, 2, \cdots)$ とする．このとき，$\theta = (\theta_1, \cdots, \theta_k)$ の次元数分だけ両方のモーメントを等置して，

$$\mu'_{r,\theta} = M'_r \quad (r = 1, \cdots, k) \tag{7・1}$$

を連立させて得られる $\theta_1, \cdots, \theta_k$ の解を $\hat{\theta}_1 = \hat{\theta}_1(\boldsymbol{X}), \cdots, \hat{\theta}_k = \hat{\theta}_k(\boldsymbol{X})$ とする．このとき，$\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$ を $\theta = (\theta_1, \cdots, \theta_k)$ の**モーメント推定量**（moment estimator）という．大数の法則から，各 $i = 1, \cdots, k$ について，連続性の仮定のもとで $\hat{\theta}_i$ は θ_i に確率収束する．いま，平均 μ，分散 σ^2 を持つ母集団分布からの無作為標本を $X_1, \cdots, X_n$ とする．このとき，$\theta = (\mu, \sigma^2)$ のモーメント推定量を求めるには，$\mu'_{1,\theta} = E_\theta(X_1) = \mu$，$\mu'_{2,\theta} = E_\theta(X_1^2) = \mu^2 + \sigma^2$ と $M'_1 = (1/n)\sum_{i=1}^n X_i = \bar{X}$，$M'_2 = (1/n)\sum_{i=1}^n X_i^2$ を等値させて，

$$\mu = \bar{X}, \quad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$$

を μ，σ^2 について解く．平均 μ，分散 σ^2 のモーメント推定量は $\hat{\theta} = (\bar{X}, S^2)$，すなわち，標本平均 $\bar{X}$ と標本分散 $S^2 = (1/n)\sum_{i=1}^n (X_i - \bar{X})^2$ になる．

■12 群 - 3 編 - 7 章

7-3 最尤法

（the method of maximum likelihood）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

確率ベクトル $\boldsymbol{X} = (X_1, \cdots, X_n)$ の j.p.d.f. または j.p.m.f. を $f_{\boldsymbol{X}}(\boldsymbol{x}, \theta)$ $(\theta \in \Theta)$ とする．ただし，$\boldsymbol{x} = (x_1, \cdots, x_n)$ とする．いま，$\boldsymbol{X}$ が実現値 $\boldsymbol{x}$ をとるとき，$L(\theta; \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}, \theta)$ とおいて，これを θ の関数とみなすとき，θ の**尤度関数**（likelihood function）という．この尤度関数 $L(\theta; \boldsymbol{x})$ は，母数の値が θ のときのデータ $\boldsymbol{x}$ の起こりやすさを表すと考えられる．そこで，これを最大にするもの，すなわち

$$\max_{\theta \in \Theta} L(\theta; \boldsymbol{x}) = L(\hat{\theta}; \boldsymbol{x}) \tag{7・1}$$

となる $\theta = \hat{\theta}(\boldsymbol{x})$ を θ の最尤推定値といい，$\hat{\theta}(\boldsymbol{X})$ を θ の**最尤推定量**（maximum likelihood estimator 略して **MLE**）という．各 $\boldsymbol{x}$ について $L(\theta; \boldsymbol{x})$ が θ に関して滑らかならば，**尤度方程式**（likelihood equation）

$$(\partial/\partial\theta) \log L(\theta; \boldsymbol{x}) = 0 \tag{7・2}$$

の解 $\theta = \hat{\theta}(\boldsymbol{x})$ から，θ の最尤推定量 $\hat{\theta}(\boldsymbol{X})$ を得ることも多い．同様に，$\theta = (\theta_1, \cdots, \theta_k) \in \Theta \subset \boldsymbol{R}^k$ についても，

$$(\partial/\partial\theta_1) \log L(\theta; \boldsymbol{x}) = 0, \cdots, (\partial/\partial\theta_k) \log L(\theta; \boldsymbol{x}) = 0 \tag{7・3}$$

の解 $\theta_1 = \hat{\theta}_1(\boldsymbol{x}), \cdots, \theta_k = \hat{\theta}_k(\boldsymbol{x})$ から，θ の最尤推定量 $\hat{\theta}(\boldsymbol{X}) = (\hat{\theta}_1(\boldsymbol{X}), \cdots, \hat{\theta}_k(\boldsymbol{X}))$ を得る．いま，母集団分布を $N(\mu, \sigma^2)$ とする．このとき，$\theta = (\mu, \sigma^2)$ の対数尤度関数は

$$\log L(\mu, \sigma^2; \boldsymbol{x}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

になるから，尤度方程式

$$\frac{\partial}{\partial\mu} \log L(\mu, \sigma^2; \boldsymbol{x}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0,$$

$$\frac{\partial}{\partial\sigma^2} \log L(\mu, \sigma^2; \boldsymbol{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

を連立させて (μ, σ^2) の解 $(\bar{x}, s^2)$ を得る．ここで，$\bar{x} = (1/n) \sum_{i=1}^{n} x_i$，$s^2 = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^2$ である．尤度関数が $\theta = (\bar{x}, s^2)$ で最大になることは，

$$2n^{-1}\{\log L(\bar{x}, s^2; \boldsymbol{x}) - \log L(\mu, \sigma^2; \boldsymbol{x})\} = \frac{s^2}{\sigma^2} - 1 - \log \frac{s^2}{\sigma^2} + \frac{(\bar{x} - \mu)^2}{\sigma^2} \geq 0$$

が任意の (μ, σ^2) について成り立つことから示される．よって，$(\bar{X}, S^2)$ は (μ, σ^2) の MLE になる．

■12 群 - 3 編 - 7 章

7-4 十分性
（sufficiency）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

確率ベクトル $\boldsymbol{X} = (X_1, ..., X_n)$ の j.p.d.f. または j.p.m.f. を $f_{\boldsymbol{X}}(\boldsymbol{x}, \theta)$ $(\theta \in \Theta)$ とする．ただし，$\boldsymbol{x} = (x_1, ..., x_n)$ とする．一般に，標本 $\boldsymbol{X}$ が持つ θ に関する情報について，適切な統計量 $T = T(\boldsymbol{X})$ は情報損失を起こさない．つまり，統計量 $T = T(\boldsymbol{X})$ について，$T = t$ が与えられたときに，$\boldsymbol{X}$ の c.p.d.f. または c.p.m.f. $f^{\theta}_{\boldsymbol{X}|T}(\boldsymbol{x}|t)$ が θ に無関係であるとき，T は θ に対する**十分統計量**（sufficient statistic）であるという．例えば，ベルヌーイ分布 Ber(θ) の θ に対して $T = \sum_{i=1}^{n} X_i$ は十分統計量になる．このことは，1 枚のコイン投げのような 2 項試行において，n 回のうち表が出る回数は T になり，表が出る確率 θ に関して $\boldsymbol{X}$ の持つ情報を T がすべて持っていることを意味する．一般に，十分統計量を見つけるためには，ネイマン (J. Neyman (1894-1981)) の**因子分解**（factorization）が有用である．統計量 $T = T(\boldsymbol{X})$ が θ に対する十分統計量であるための必要十分条件は，$\boldsymbol{X}$ の j.p.d.f. または j.p.m.f. $f_{\boldsymbol{X}}$ が

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \theta) = g_{\theta}(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{7・1}$$

の形に分解できることである．この命題を用いれば，$\boldsymbol{X}$ の分布が指数型分布族に属していれば十分統計量を見つけることが容易になる．一般に，多くの十分統計量が存在するが，$\boldsymbol{X}$ の最大縮約を持つ十分統計量（これを**最小十分統計量**（minimal sufficient statistic）という）が望ましい．また，統計量 T について T の関数 h が，任意の $\theta \in \Theta$ について $E_{\theta}[h(T)] = 0$ ならば $h(T) \equiv 0$ が成り立つとき，T は θ に対して**完備**（complete）であるという．

■12 群 - 3 編 - 7 章

7-5 一様最小分散不偏推定

（uniformly minimum variance unbiased estimate）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

母数 θ を持つ母集団分布からの無作為標本を $X_1, \cdots, X_n$ とする．ただし，$\theta \in \Theta$ とする．ここで，$\boldsymbol{X} = (X_1, \cdots, X_n)$ に基づく θ の関数 $g(\theta)$ の推定量 $\hat{g}_n = \hat{g}_n(\boldsymbol{X})$ について，任意の $\theta \in \Theta$ に対して，$E_\theta[\hat{g}_n(\boldsymbol{X})] = g(\theta) + b_n(\theta)$ とするとき，$b_n(\theta)$ を $\hat{g}_n$ の $g(\theta)$ に対する**偏り**（bias）という．偏り $b(\theta) \equiv 0$ となるとき，すなわち $g(\theta)$ の推定量 $\hat{g}_n = \hat{g}_n(\boldsymbol{X})$ が，任意の $\theta \in \Theta$ について $E_\theta[\hat{g}_n(\boldsymbol{X})] = g(\theta)$ であるとき，$\hat{g}_n$ を $g(\theta)$ の**不偏推定量**（unbiased estimator）という．推定量のクラスを不偏推定量全体に制限して，その中で分散を最小にする推定量を求めることは，推定量の良さを測る尺度の一つとしてよく知られている．いま，$g(\theta)$ の不偏推定量のクラスを $\mathcal{U}$ とする．任意の $\theta \in \Theta$ に対して，

$$\min_{\hat{g}_n \in \mathcal{U}} V_\theta(\hat{g}_n) = V_\theta(\hat{g}_n^\star) \tag{7・1}$$

となる $g(\theta)$ の不偏推定量 $\hat{g}_n^\star = \hat{g}_n^\star(\boldsymbol{X})$ が存在するとき，$\hat{g}_n^\star$ を**一様最小分散不偏**（uniformly minimum variance unbiased 略して **UMVU**）**推定量**という．UMVU 推定量を見つける方法として，**クラメール・ラオの不等式**などの情報不等式の下界を用いる方法や，完備な十分統計量に基づいた不偏推定量を用いる方法などが知られている．

■12 群 - 3 編 - 7 章

7-6 区間推定

（interval estimation）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

母数 θ を持つ母集団分布からの無作為標本を $X_1, \cdots, X_n$ とする．ただし $\theta \in \Theta$ とし，Θ を $\boldsymbol{R}^1$ の開区間とする．いま，$0 < \alpha < 1$ とするとき，$\boldsymbol{X} = (X_1, \cdots, X_n)$ に対して Θ の閉区間 $[a(\boldsymbol{X}),\ b(\boldsymbol{X})]$ を定めて，任意の $\theta \in \Theta$ について

$$P^{\theta}_{\boldsymbol{X}}\{a(\boldsymbol{X}) \leq \theta \leq b(\boldsymbol{X})\} \geq 1 - \alpha \tag{7・1}$$

とする．このとき, $[a(\boldsymbol{X}),\ b(\boldsymbol{X})]$ を**信頼係数**（confidence coefficient）（または**信頼度**（confidence level））$1-\alpha$ の θ の**信頼区間**（confidence interval）といい，$a(\boldsymbol{X}),\ b(\boldsymbol{X})$ を**信頼限界**（confidence limits）という．式 (6·1) の確率の値に等号が成り立つとき，信頼区間は**相似**（similar）であるという．また，$\boldsymbol{X}$ の関数 $\underline{\theta}(\boldsymbol{X}),\ \overline{\theta}(\boldsymbol{X})$ を定めて，任意の $\theta \in \Theta$ について

$$P^{\theta}_{\boldsymbol{X}}\{\underline{\theta}(\boldsymbol{X}) \leq \theta\} = 1 - \alpha, \quad P^{\theta}_{\boldsymbol{X}}\{\overline{\theta}(\boldsymbol{X}) \geq \theta\} = 1 - \alpha \tag{7・2}$$

とするとき，$\underline{\theta}(\boldsymbol{X}),\ \overline{\theta}(\boldsymbol{X})$ をそれぞれ信頼係数 $1-\alpha$ の θ の**下側信頼限界**（lower confidence limit），**上側信頼限界**（upper confidence limit）という．更に，$\boldsymbol{X}$ の実現値 $\boldsymbol{x} = (x_1, \cdots .x_n)$ について，区間 $[a(\boldsymbol{x}),\ b(\boldsymbol{x})]$ を θ の $100(1-\alpha)$%信頼区間という．通常，α としては，0.05 をとるが，0.10，0.01 などをとることもある．上のような信頼区間で，母数 θ を推定することを**区間推定**（interval estimation）という．実は，区間推定は仮説検定と密接な関係を持ち，信頼区間は両側検定の受容域に対応している．信頼区間を構成するには，次の量を考える．θ に依存する $\boldsymbol{X}$ の実数値（可測）関数 $T(\boldsymbol{X}, \theta)$ の分布が θ に無関係になるとき，$T(\boldsymbol{X}, \theta)$ を**枢軸量**（pivotal quantity）という．適当な枢軸量 $T(\boldsymbol{X}, \theta)$ を見つけて，任意の α $(0 < \alpha < 1)$ について，ある閉区間 $[t_1, t_2]$ をとって

$$P^{\theta}_{\boldsymbol{X}}\{t_1 \leq T(\boldsymbol{X}, \theta) \leq t_2\} = 1 - \alpha \tag{7・3}$$

とできて，更に $\{\theta | t_1 \leq T(\boldsymbol{X}, \theta) \leq t_2\}$ が閉区間 $[\underline{\theta},\ \overline{\theta}]$ になれば，これが信頼係数 $1-\alpha$ の θ の信頼区間になる．ここで，$T(\boldsymbol{X}, \theta)$ の分布が θ に無関係なので，t_1, t_2 は θ に無関係に定まる．

■12 群 - 3 編 - 7 章

7-7 仮説検定

（hypothesis testing）

（執筆者：青嶋　誠）［2009 年 2 月 受領］

母集団から抽出された標本に基づいて，母集団に関する何らかの結論を導き出したい．母集団分布がある母数 θ を持っている場合に，例えば $H:\theta=\theta_0$ という仮説を設けて，これが正しいか否かを標本に基づいて判定することを**仮説検定**（hypothesis testing）という．母集団分布から得られた無作為標本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ に基づいて適当な領域 $R_{\boldsymbol{X}}$ を定めて，$\boldsymbol{X}$ の実現値 $\boldsymbol{x}=(x_1,\cdots,x_n)$ について，$\boldsymbol{x}\in R_{\boldsymbol{X}}$ のとき H を棄却し，$\boldsymbol{x}\notin R_{\boldsymbol{X}}$ のとき H を受容するという決定を行う．ここで，調べたい仮説 H を**帰無仮説**（null hypothesis）といい，この仮説と比較するための仮説 K を**対立仮説**（alternative hypothesis）という．$R_{\boldsymbol{X}}$ を**棄却域**（rejection region）といい，その補集合を**受容域**（acceptance region）という．検定の実施には，2 種類の過誤をおかすことは避けられない．一つは，H が真（K が偽）のときに H を棄却（K を受容）する**第 1 種の過誤**（error of type I）と呼ばれるもの，いま一つは，H が偽（K が真）のときに H を受容（K を棄却）する**第 2 種の過誤**（error of type II）と呼ばれるものである．これら 2 つの過誤の確率を同時に小さくするような検定を行いたい，すなわち棄却域を選びたいが，これは無理である．そこで，あらかじめ定められた $\alpha\in(0,1)$ に対して第 1 種の過誤の確率 α_R を α 以下にするという条件のもとで，第 2 種の過誤の確率 β_R を最小にするように棄却域を選ぶようにする．ここで，α を**有意水準**（level of significance）といい，通常，0.05, 0.01, 0.1 の値をとることが多い．また，$\gamma_R=1-\beta_R$ とおいて，これを**検出力**（power）といい，対立仮説 K が真であるとき仮説 H を棄却する確率を表す．第 2 種の過誤の確率 β_R を最小にすることは，検出力を最大にすることと同値である．仮説検定と区間推定は密接な関係があり，$H:\theta=\theta_0$，$K:\theta\neq\theta_0$ に関する検定（これを**両側検定**（two-sided test）という）を行ったとき，受容域は信頼係数 $1-\alpha$ の信頼区間に対応している．したがって，信頼区間は理論的には検定の受容域から導出され，検定の良さの基準はそのまま信頼区間にも適応され得る．

■12 群 - 3 編 - 7 章

7-8 統計モデル

（statistical models）

（執筆者：青嶋　誠）[2009 年 2 月 受領]

確率変数 X_i が p.d.f. $p(x-\mu)$ $(i = 1, \cdots, n)$ を持ち，各 X_i は互いに独立である 1 標本のモデルを考える．いま，$p(x)$ は 0 について対称，すなわち $p(x) = p(-x)$ を仮定する．そのとき，平均は $E(X_i) = \int_{-\infty}^{\infty} xp(x-\mu)dx = \mu$ となる．仮説検定の帰無仮説として $H: \mu = \mu_0$ を考える．平均 μ の点推定は標本平均 $\hat{\mu} = \overline{X}$ で与えられ，これに基づく t 検定は母集団分布が正規分布に従うとき最良な手法になる．これを**パラメトリック論**（parametric theory）による手法という．パラメトリック論による手法には，幾つかの欠点がある．第一に，データの中に異常値（外れ値）を含む場合に，推定結果や検定結果が異常値に振り回されやすい．第二に，$p(x)$ が裾の重い分布のときには推測法として良くない．第三に，標本サイズが大きくないとき，$p(x)$ が既知でない限り検定手法を使うことは難しい．これに対して，$p(x)$ が未知でも推定や検定が行える手法を**ノンパラメトリック論**（nonparametric theory）による手法という．順位統計量 T の値が大きいならば H を棄却する，という順位検定はよく知られる．$p(x)$ が正規分布のときは順位検定は t 検定よりも少し劣るが，$p(x)$ が正規分布から離れた分布のときには t 検定よりも検出力が良いことが知られている．特に，裾の重い分布に対して，順位検定は有効である．すなわち，解析結果が異常値にあまり影響を受けない．ノンパラメトリックとパラメトリックの中間的な立場に，**セミパラメトリック論**（semiparametric theory）による手法がある．これは，$p(x)$ が正規分布の ε 近傍 $U_\varepsilon = \{p(x) = (1-\varepsilon)\phi(x) + \varepsilon h(x):$ $\phi(x)$ は標準正規分布の密度関数，$h(x)$ はすべての x に対して $h(-x) = h(x)$ を満たすある密度関数 $\}$ の中にあるとき，近似的に最良な手法になる．

■参考文献

1)　赤平昌文 :“ 統計解析入門, ” 森北出版，2003.

2)　稲垣宣生 :“ 数理統計学（改訂版）, ” 裳華房，2003.

3)　白石高章 :“ 統計科学，日本評論社，2003.

4)　竹内　啓, 他編 :“ 統計学辞典, ” 東洋経済新報社，1989.